图书情报工作
第65卷 第9期 2021年5月



# 泄露概率情境下的个人数据隐私计量研究*

■ 张凯亮[1] 臧国全[2,3]
[1] 郑州大学政治与公共管理学院 郑州 450001 [2] 郑州大学信息管理学院 郑州 450001
[3] 郑州市数据科学研究中心 郑州 450001

摘 要：[目的/意义]与确定泄露的情境相比，个人数据以一定概率泄露的情境更为普遍。由此计量泄露概率情境下的用户个人隐私价值，提出隐私计量的新视角，计量结果对隐私分级保护也具有实际意义。[方法/过程]基于多级价格表的方法，计量用户的金融风险回报率；基于用户的金融风险回报率，改造多级价格表的实现机制，引导用户在无隐私泄露概率方案和存在隐私泄露概率的方案中做出决策，测度用户在一定泄露概率情境下对其个人数据的价值认知。[结果/结论]在泄露概率为30%的情境下，用户对其社交网络中的个人数据隐私价值认知约为89.5元；同时，在泄露概率为100%的情境下以接受意愿和支付意愿体现的个人数据隐私价值分别为124.1和93.8元。表明用户在泄露概率情境下对其个人数据隐私价值的认知取决于个人数据隐私价值本身和用户对泄露概率风险容忍度两个方面。

关键词：个人隐私 泄露概率 隐私价值 多级价格表机制
分类号：G251
DOI：10.13266/j.issn.0252-3116.2021.09.007



## 1 引言

大数据背景下，个人数据隐私问题吸引了广泛的社会关注。无论是政府公共部门还是社会企业组织都意识到海量个人数据蕴含巨大价值，比如公共部门通过收集、整合个人数据有助于提高社会服务效率；企业组织通过挖掘、分析个人数据有助于提供个人定制服务等。然而，“华住酒店1.3亿用户个人数据泄露”[1]、“领英1.59亿用户敏感数据泄露”[2]等层出不穷的个人数据泄露事件严重威胁到了社会公众的切身利益安全，公众逐渐意识到保护个人数据隐私的必要性。

用户在披露个人数据后带来的便捷服务和隐私风险之间做抉择时，会权衡成本与收益，成本如隐私泄露、价格歧视、身份盗窃等，收益如个性化服务、价格折扣等。企业在收集、使用用户数据时，需权衡用户隐私保护支出与企业收入提升之间的关系。平衡隐私关注是企业和用户两者都面临的一个棘手问题，解决这个问题的关键是计量用户个人数据的隐私价值。计量用户个人数据的隐私价值一方面能够帮助企业制定差异化隐私保护政策，提高隐私保护效率，降低隐私保护成本；另一方面，能够为公共部门制定用户个人数据隐私泄露的赔偿标准提供参考。

个人数据存在泄露概率，目前的研究均集中在个人数据确定泄露的情境，但客观上，泄露是一个概率性事件。因此，计量在一定泄露概率条件下，用户对其个人数据隐私价值的认知也具有实际意义。

## 2 个人数据隐私价值计量相关研究

已有的个人数据隐私价值计量相关研究汇总见表1。

根据表1，已有的实证研究主要计量个人数据隐私的货币价值，表现形式有两种：①用户愿意为保护其个人数据隐私免遭侵犯而支付的货币价格(Willing to Pay, WTP)。如，S. Egelman等[17]通过离散选择分析法计量得出用户位置数据和通话记录的WTP约为1.5美元；J. Kim等[13]通过联合分析法计量得出韩国公众平均每月愿意花费约6.8美元用于个人信息保护服务。②用户愿意牺牲其个人数据隐私而能够接受的货币补偿价格(Willing to Accept, WTA)。如，臧国全等[11]

* 本文系国家自然科学基金项目“数字保存的风险型元数据与风险监控研究”(项目编号:71673255)研究成果之一。
作者简介：张凯亮(ORCID:0000-0003-0454-0525)，博士研究生，E-mail:779015223@qq.com；臧国全(ORCID:0000-0002-9606-6455)，院长，教授，博士生导师。
收稿日期:2020-12-08 本文起止页码:62-69 本文责任编辑:易飞

**表 1　已有的个人数据隐私价值计量相关研究**

| 价值表现 | 计量方法 | 作者(发表年) |
|---|---|---|
| 支付意愿 | 联合分析法 | I. H. Hann 等[3](2002);A. Acquisti 等[4](2012) |
| | 离散选择分析法 | D. Potoglou 等[5](2013);S. Lim 等[6](2018);邓胜利等[7](2019) |
| | 条件价值法 | J. Grossklags 等[8](2007);黄逸珺等[9](2017) |
| | 拍卖实验法 | B. A. Huberman 等[10](2005);臧国全等[11](2020) |
| 接受意愿 | 联合分析法 | H. Krasnova 等[12](2009);J. Kim 等[13](2015);Y. Pu 等[14](2016) |
| | 离散选择分析法 | J. Y. Tsai 等[15](2011);A. R. Beresford 等[16](2012);S. Egelman 等[17](2013);H. Krasnova 等[18](2014) |
| | 条件价值法 | J. E. Kim 等[19](2010);M. Otsuki 等[20](2013) |
| | 拍卖实验法 | G. Danezis 等[21](2005);S. Spiekermann 等[22](2012) |

基于改造过的 BDM 机制计量得出用户个人偏好数据、联系资料数据的 WTA 分别约为 38.8 元、136.3 元;黄逸珺等[9]采用条件价值法调查得出用户年龄数据和家庭地址的 WTA 分别约为 6.99 元、66.19 元。

根据表 1,已有的个人数据隐私价值计量方法可分为问卷调查法和拍卖实验法两类:①问卷调查法包括联合分析法(Conjoint Analysis)、离散选择分析法(Discrete Choice Analysis)和条件价值法(Contingent Valuation Method)。其中,联合分析法和离散选择分析法均是根据产品属性及属性水平构建产品轮廓,采用正交法形成一定数量的产品集,但二者的评价方式不同,联合分析法采用打分、排序等方法,而离散选择分析法是让用户实际选择何种产品或服务。条件价值法是通过假设情景,直接询问用户对其个人数据隐私的 WTP 或 WTA。②拍卖实验法包括 $n$ 级价格拍卖法($n-1$ 个最高竞价者获胜,并支付第 $n$ 个最高价)、Becker DeGroot Marschak(BDM)拍卖法(竞价者与随机函数竞价,竞价者价格高于随机函数返回的价格时,竞价者获胜,并以随机随机函数返回的价格支付)和随机 N 级价格拍卖法(竞价者 $i$ 是与其他 $N-1$ 位参与者竞价,随机抽取 N 位竞拍者中的一个出价作为市场出清价格,出价大于或等于该价格的竞价者获胜,并以该价格支付)。其中,$n$ 级价格拍卖法和随机 $N$ 级价格拍卖法适用于同类多件物品拍卖,BDM 拍卖法适用于单件物品拍卖。

由上可知,4 种主要计量方法中,每种方法都可以从支付意愿和接受意愿两个维度计量个人数据隐私价值。尽管计量结果可能存在差异,但不影响个人数据隐私价值大小的相对性特质。

全面考察上述隐私价值测量的问卷调查法和拍卖实验法的实证研究后,作者发现已有研究均假定个人隐私确定泄露,没有考虑仅以一定概率泄露的情况。实际上,个人隐私决策不仅取决于个人数据隐私本身价值,还包括泄露概率。完全确定泄露与一定可能性泄露导致的个人隐私关注程度可能存在差异,比如,个人数据保存系统中的隐私保护措施就是为了降低隐私泄露概率从而减轻用户隐私担忧。因此,本研究拟测度用户在泄露概率情境下对其个人数据隐私的价值认知,简称泄露概率隐私价值(Value of Privacy under Leakage Probability,VPLP)。

## 3　计量模型

风险的涵义是不期望事件的发生概率[23]。因此,风险的构成有两个要素:不期望事件、发生概率。个人数据的隐私泄露是一个风险事件,因为,个人数据的泄露是数据主体不期望发生的事件,同时,泄露本身又是一个概率性事件。在个人数据保存系统中,个人数据可能被泄露,也可能不被泄露,取决于安全保护措施,保护措施越严格,泄露概率越低,反之,泄露概率越高。因此,本文使用"隐私泄露风险"来显性地表达"隐私泄露这类事件的发生概率"。

本研究采用多级价格表机制(Multiple Price List,MPL),设计易于理解的金融概率风险多级价格表,计量用户金融风险回报率(Return on Financial Risk,RFR)。基于用户金融风险回报率的计量结果,测度用户在泄露概率情境下的个人数据隐私价值认知。

### 3.1　金融风险回报率

MPL 机制是实验经济学领域测量用户风险态度的一种常用手段,最早由 H. P. Binswanger[24]提出,通过向用户展示一系列不同收益的方案组合,每个组合中均有一个安全方案(不存在损失的概率)和风险方案(存在损失的概率),安全方案的固定收益逐行递减,而所有风险方案的期望收益相同,询问用户在安全方案的固定收益降低至哪个水平时,会放弃安全收益而选择追求可能更高的风险收益,通过观察用户决策间接推理出用户的风险态度。采用 MPL 机制,设计 RFR 多级价格表,如表 2 所示:

表2 RFR多级价格

| 行号 | 方案A | 方案B |
|---|---|---|
| 0 | 您可获得 $x$ 元 | 您可获得 $y$ 元,但有 $p$ 的概率损失 $c$ 元,即您有 $(1-p)$ 的概率获得 $y$ 元,有 $p$ 的概率获得 $(y-c)$ 元 |
| 1 | 您可获得 $(x-a)$ 元 | 您可获得 $y$ 元,但有 $p$ 的概率损失 $c$ 元,即您有 $(1-p)$ 的概率获得 $y$ 元,有 $p$ 的概率获得 $(y-c)$ 元 |
| 2 | 您可获得 $(x-2a)$ 元 | 您可获得 $y$ 元,但有 $p$ 的概率损失 $c$ 元,即您有 $(1-p)$ 的概率获得 $y$ 元,有 $p$ 的概率获得 $(y-c)$ 元 |
| … | … | … |
| $k$ | 您可获得 $(x-k*a)$ 元 | 您可获得 $y$ 元,但有 $p$ 的概率损失 $c$ 元,即您有 $(1-p)$ 的概率获得 $y$ 元,有 $p$ 的概率获得 $(y-c)$ 元 |

表2包含两个方案:方案A为安全方案,收益以 $a$ 元逐行递减,在第 $i$ 行,方案A的收益为 $Y(A_i)=x-i*a, i\in[0,k]$;方案B为风险方案(有 $p$ 的概率损失 $c$ 元),每行方案B期望收益为 $E(B_i)=y*(1-p)+(y-c)*p$。当用户决策在第 $i$ 行由方案A转向方案B时,代表用户舍弃安全收益而选择风险收益,此时该用户的RFR为:

$$RFR=\frac{E(B_i)-Y(A_i)}{Y(A_i)}, i\in[0,k]$$

### 3.2 泄露概率隐私价值计量

通过将RFR多级价格表(见表2)中方案B的"有 $p$ 的概率损失 $c$ 元"改变为"有 $p$ 的概率个人隐私被泄露",可以模拟出真实的泄露概率情境,从而形成VPLP多级价格表(见表3)。隐私价值是隐私主体对隐私客体遭受侵犯时可能导致的损失估计,这种损失包含人格和财产两个方面(经济社会的发展导致人格损失的财产化转换越来越普遍),可用货币价格表示用户对这类隐私价值的认知。因此,在VPLP多级价格表中方案B可货币化表示为"您可获得 $y$ 元,但有 $p$ 的概率损失 $vplp$ 元"。

表3 VPLP多级价格

| 行号 | 方案A | 方案B |
|---|---|---|
| 0 | 您可获得 $x$ 元 | 您可获得 $y$ 元,但有 $p$ 的概率泄露您的个人隐私,即,您可获得 $y$ 元,但有 $p$ 的概率损失 $vplp$ 元 |
| 1 | 您可获得 $(x-a)$ 元 | 您可获得 $y$ 元,但有 $p$ 的概率泄露您的个人隐私,即,您可获得 $y$ 元,但有 $p$ 的概率损失 $vplp$ 元 |
| 2 | 您可获得 $(x-2a)$ 元 | 您可获得 $y$ 元,但有 $p$ 的概率泄露您的个人隐私,即,您可获得 $y$ 元,但有 $p$ 的概率损失 $vplp$ 元 |
| … | … | … |
| $k$ | 您可获得 $(x-k*a)$ 元 | 您可获得 $y$ 元,但有 $p$ 的概率泄露您的个人隐私,即,您可获得 $y$ 元,但有 $p$ 的概率损失 $vplp$ 元 |

基于表3的设计,将用户的隐私决策转换为在金融情境下安全方案和风险方案之间的决策。实际上,表2情境属于直接金融情境(所有金融要素直接采用货币形式表达),表3情境属于间接金融情境(VPLP以用户价值认知方式间接采用货币表达)。两者均属于金融领域,当表2、表3情境采用相同的损失概率时,用户在表2和表3情境下的金融风险回报率应该相同。由此推导出VPLP的计量公式为:

$$VPLP=\frac{y-(RFR+1)*(x-i*a)}{p}, i\in[0,k]$$

其中,$i$ 为用户在表3中的选择由方案A转向方案B的行号,RFR为基于表2计算出的该用户金融概率风险回报率。

## 4 计量实验

### 4.1 个人数据样本

本实验通过调查问卷的方式获取个人基本数据和个人隐私数据样本。问卷应包含3部分内容:

#### 4.1.1 人口统计学数据

包括用户姓名、性别、年龄、学历等个人标识和半标识数据,用于基于人口统计特征的泄露概率隐私值分析。

#### 4.1.2 个人隐私数据

从计量模型可以看出,该模型适合于任何领域的个人隐私数据,可以是宏观上的隐私数据条目,也可以是微观上的隐私数据项目。由于实验过程中,可能需要用户公开个人隐私数据,因此,调查中用户个人隐私数据需要能够及时现场提供。基于该要求,本实验选择用户的微信数据(包括微信账户页面截图、近半年已发布朋友圈页面截图、聊天主界面截图等)和微博数据(包括微博账户页面截图、近半年已发布微博页面截图、近半年微博评论界面截图等)作为用户个人隐私数据。

#### 4.1.3 个人隐私价值的直接认知数据

目的是获取用户对其个人隐私价值认知的直接报告值,包括用户为保护其个人隐私数据而愿意支付的价格(WTP)和用户披露个人隐私数据时能够接受的补偿价格(WTA)。

### 4.2 实验步骤

#### 4.2.1 步骤1:计量RFR

首先,实验者向用户发放一张RFR多级价格表,该表包含 $K+1$ 行,每行包括安全方案A和风险方案B,用户需在每一行两个方案之间做出一个选择。其次,记录用户决策从方案A转至方案B的初始行号,计算该用户的RFR。最后,采用随机函数返回一个0-k的行号i,以用户在该行的选择向用户支付实验报酬。若用户在该行选择方案A,则向用户支付 $(x-i*$

$a$)元。若在该行选择方案 B,再采用[0,1]随机函数确定方案 B 的执行结果,若随机函数返回值大于 $p$,则向用户支付 $y$ 元;反之,支付($y-c$)元(实验者已在实验协议中告知用户,基于实验成本考虑,步骤 1 和 2 均以 50:1 的比例向用户支付实际金额。)。

在 RFR 实验中,采用随机函数确定行号,并以用户在该行的选择向用户支付实验报酬,目的是保证用户决策的真诚性。该方法源自于 Becker DeGroot Marschak(BDM)拍卖机制[25]的思想,若用户的实际选择与真实想法不同,随机函数返回行号 $i$ 落在用户的说谎区间,用户需承担方案 A、B 之间的预期收益差。只有当用户实际选择与真实想法一致时,不需要承担额外风险。因此,该方法能够测量用户真实的 RFR。同理,在下述步骤 2 中,同样采用该方法。

#### 4.2.2 步骤 2:计量 VPLP

首先,实验者向用户发放一张 VPLP 多级价格表,该表包含 K+1 行,每行包括安全方案 A 和风险方案 B,用户需在每一行两个方案之间做出一个选择。其次,记录用户决策从方案 A 转至方案 B 的初始行号,采用该用户的 RFR,计算其 VPLP。最后,采用随机函数返回一个 0－k 的行号 $i$,以用户在该行的选择向用户支付实验报酬。若用户在该行选择方案 A,则向用户支付($x-i*a$)元。若在该行选择方案 B,再采用[0,1]随机函数确定方案 B 的执行结果。若随机函数返回值大于 $p$,向用户支付 $y$ 元,且无需披露用户个人隐私数据;反之,向用户支付 $y$ 元,但需向实验群组内所有用户披露其个人隐私数据(既可模拟隐私泄露的真实情景,也可控制隐私泄露的范围,以避免真实地泄露用户个人隐私)。

### 4.3 预实验

依据正式实验的目标用户数量,根据样本抽样数量的经验法则[26],计算预实验的用户数量。此次仅进行 VPLP 计量预实验,预实验目的是确定相关参数。这些参数确定后,用于 RFR 和 VPLP 正式实验。

#### 4.3.1 参数 *x* 的确定

参数 $x$ 的取值应大于或等于参数 y,以保证所有用户决策均存在从方案 A 到方案 B 的变化,增加实验结果的可信度,否则可能出现风险厌恶度较低且对个人隐私价值认知较低的用户在所有行均选择方案 B 的现象,这种现象无法计量金融风险回报率,导致也无法计量泄露概率隐私值。当然,最简单的方法是将参数 $x$ 的取值设置与参数 $y$ 相等。

#### 4.3.2 参数 *y* 的确定

根据前景理论[27],当两个方案的期望收益相等时,若其中一个方案存在用户收益为负值的情况,用户对该风险(收益为负)的厌恶会降低选择该方案的意愿。因此,参数 $y$ 的取值应大于 VPLP 的预估值,以避免用户选择方案 B 获取的货币收益不足以弥补披露个人隐私数据带来的损失,从而降低其选择方案 B 的意愿。作者在以往的研究中,采用 BDM 拍卖机制,测量出用户网络隐私数据的 WTA 约在 38.8－237.9 元之间[11]。因此,本实验基于收集的个人隐私样本,将预实验的参数 $y$ 设置为 150 元。可通过预实验结果验证该参数的取值是否合理,若预实验中大多数用户均未选择过方案 B,或仅在极少行选择方案 B,说明参数 $y$ 的取值过低,应设置更高的 $y$ 值;反之,则说明参数 y 的取值合理。基于预实验结果,$y=150$ 时,96.7% 的用户选择过方案 B 且选择方案 B 的平均行数为 13 行,故将参数 $y$ 的取值设置为 150 元是合理的。

#### 4.3.3 参数 *a*、*k* 的确定

实际上,参数 $a$ 仅反映方案 A 收益的递减速率,其取值并不会对实验结果产生明显影响。但参数 $a$ 的取值也不能过大或过小,参数 $a$ 过大导致相邻两行中方案 A 的收益相差过多,可能无法准确观察用户的决策行为;参数 $a$ 过小导致多级价格表的行数过多,可能影响用户的决策体验。因此,本文基于参数 $y$ 的初始值,将参数 $a$ 的取值设置为 5。参数 k 决定多级价格表的行数,只有当多级价格表最后一行中方案 A 的期望收益远小于方案 B 时,才可能存在用户决策从方案 A 转向方案 B。因此,基于参数 $a$ 的取值,本文将参数 k 的取值设置为 19(总行数为 k+1 行,共 20 行),即在多级价格表最后一行中方案 A 的收益为 55 元。

#### 4.3.4 概率 *p* 的确定

设定概率 $p$ 时应考虑现实情况下的隐私泄露状况,最理想的方式是统计全社会的个人隐私数据总量和被泄露的个人隐私数据量,据此计算现实情况下的隐私泄露概率。然而,作者经过多方检索均未发现相关的权威统计数据。作为一种折衷方案,本文选择预实验的方式,选取部分用户,观察其决策结果(用户决策是基于其自身经历、感知所做出的判断,具备可信度),从而确定概率 $p$ 的取值。研究设置 4 组对照预实验,将泄露概率 $p$ 分别设定为 20%、30%、40%、50%(过高或过低的泄露概率可能不太符合实际),各组预实验的其他参数相同。基于预实验的结果分布,$p=30\%$ 时,用户在 VPLP 多级价格表中转换行号的分布

趋于正态分布，故选择概率 $p$ 的取值为 30% 较合理。

### 4.4 正式实验

同一用户需分别完成 RFR 计量正式实验和 VPLP 计量正式实验。所有正式实验均通过自编软件实施，自编软件通过 Unity 工具和 C#语言实现，软件示例如图 1 所示：

图 1 实验软件截图

设计软件时，可通过相关操作避免用户的异常实验行为。首先，当用户在多级价格表的所有行均选择方案 B 时，系统自动拒绝该用户提交实验结果；其次，用户在进行决策时，只需点击在某一行选择方案 B 即可，该行以上自动选择方案 A，该行及以下自动选择方案 B。通过上述操作设定，一方面可节省用户实验时间，提高实验效率；另一方面可避免出现异常数据。

## 5 样本统计与数据分析

### 5.1 样本统计

本实验于 2020 年 8 月实施，作者与郑州大学情报学硕士研究生共 10 人担任实验员，预实验采用现场实验方式，正式实验采用网络群组和邮件相结合的方式进行。由于实验涉及到风险容忍度，还需要在两个表格上判断选择，比较复杂，受访者需具备一定的知识水平，因此，本次实验选择郑州大学在校大学生和研究生为实验对象。每位用户在参与实验前，需仔细阅读实验步骤，理解实验内容，同意参与实验的用户需与实验组签订实验协议，且可获得 5 元报酬。

本实验共有 275 人参与，用户样本的描述性统计结果见表 4，实验结果见表 5。

表 4 描述性统计结果

| 人口统计学特征 | 类别 | 人数/人 | 频率/% |
|---|---|---|---|
| 性别 | 男 | 133 | 48.4 |
| | 女 | 142 | 51.6 |
| 学历 | 本科 | 108 | 39.3 |
| | 硕士 | 103 | 37.5 |
| | 博士 | 64 | 23.2 |

表 5 实验结果

| 类别 | 最小值 | 最大值 | 平均值 | 标准差 |
|---|---|---|---|---|
| RFR | -0.14 | 1.18 | 0.27 | 0.25 |
| VPLP | 11 | 195 | 89.5 | 36.7 |

（1）用户的 RFR。用户的 RFR 是基于用户在 RFR 多级价格表中的选择开始从安全方案 A 转到风险方案 B 时，方案 A 的安全固定收益和方案 B 的期望收益计算的。根据 RFR 的计算公式，若 RFR > 0，表明当风险方案 B 的期望收益大于安全方案 A 的固定收益时，用户才会选择风险方案 B，说明该用户不愿意承担风险方案 B 的期望收益低于安全方案 A 的固定收益时的预期收益差，即该用户是风险厌恶者；同样道理，若 RFR < 0，则说明该用户是风险偏好者；若 RFR = 0，则说明该用户是风险中性者。因此，RFR 计量实验结果表明：12% 用户为风险中性者（RFR = 0），这部分用户仅根据期望收益进行决策。82.5% 用户为风险厌恶者（RFR > 0），其中有 3.3% 用户为风险极度厌恶者（RFR = 1.18，从初始行开始就一直选择方案 A）。5.5% 用户为风险偏好者（RFR < 0），但并没有用户从初始行开始就一直选择方案 B。没有用户从初始行开始就一直选择方案 B 的原因是初始行中风险方案 B 的最大收益不高于方案 A 的安全收益，用户没必要承担方案 B 的风险，即从初始行开始就选择方案 B 的行为不合理（设计实验软件时，已将此情况进行规避）。

（2）用户的 VPLP。实验结果表明，RFR 相同的用户，其 VPLP 可能存在不同；RFR 较低的用户，也可能表现出较高的 VPLP；同样，RFR 较高的用户也可能表

张凯亮，臧国全. 泄露概率情境下的个人数据隐私计量研究[J]. 图书情报工作,2021,65(9):62－69.

现出较低的 VPLP;WTA 或 WTP 相同的用户,其 VPLP 也可能存在差异。综合说明,用户的 VPLP 取决于个人隐私价值本身(在个人隐私确定被泄露的情况下)和泄露概率两方面的因素。

### 5.2 VPLP 与人口统计学特征的相关分析

通过 K-S 正态检验,VPLP 的值符合正态分布(K-S 正态检验,$p=0.018$),因此对 VPLP 与人口统计学特征进行相关性分析时,可采用独立样本 T 检验。

#### 5.2.1 性别方面

男性用户的 VPLP 均值为 81.1,女性为 97.5。以用户性别为分组变量,对 VPLP 的值进行独立样本 T 检验。结果表明,女性用户的 VPLP 显著高于男性(独立样本 T 检验,$p<0.001$)。用户风险态度领域的研究表明,女性对风险的厌恶程度显著高于男性[28]。隐私计量领域,已有研究表明女性对其个人隐私的价值认知显著高于男性[7,19]。而 VPLP 是用户对泄露概率和个人隐私价值的综合认知,因此,女性用户的 VPLP 值高于男性是合理的。

年龄方面。由于本实验整体样本的年龄跨度不是很大,无法以年龄区间进行分组。本文仅选择部分有代表性(存在一定年龄差距)的用户进行检验,比如 27－28 岁学生的 VPLP 显著高于 18－19 岁的学生(独立样本 T 检验,$p=0.014$)。已有研究表明,用户随年龄增长,其风险厌恶程度会逐渐提高[28]。在今后的研究中,应扩大样本数量以分析用户年龄对 VPLP 的影响。

#### 5.2.2 学历方面

本科生、硕士生和博士生的 VPLP 均值分别为 83.3、91.6 和 96.8。以学历为分组变量,两两分组,分别进行独立样本 T 检验。检验结果见表 6。

表 6　独立样本 T 检验结果

| 类别 | 独立样本 T 检验 |
|---|---|
| 本科生 VS. 硕士生 | $P=0.115$ |
| 硕士生 VS. 博士生 | $P=0.392$ |
| 本科生 VS. 博士生 | $P=0.011$ |

根据表 6,博士生的 VPLP 高于硕士生,硕士生的 VPLP 高于本科生,但不存在显著性,而博士生的 VPLP 显著高于本科生。这可能和本文选取的用户样本有关,因为本文选取的用户样本均为高校在校学生,每个学历层次之间的学生年龄跨度并不是很大且存在年龄交叉,比如存在部分硕士生年龄小于本科生、部分博士生年龄小于硕士生等现象。实际上,已有研究表明,受教育程度会显著影响用户对其个人隐私的价值认知[9,20]。在今后的研究中,应扩大样本数量以分析用户学历对 VPLP 的影响。

### 5.3 VPLP 与问卷调查数据的相关分析

VPLP 与 WTP、WTA 的相关性分析。VPLP、WTP 和 WTA 的均值分别为 89.5、93.8、124.1。采用配对样本 T 检验,将 VPLP 与 WTP、WTA 分别进行相关性分析。检验结果表明,WTA 显著高于 WTP(配对样本 T 检验,$p<0.001$),WTP 显著高于 VPLP(配对样本 T 检验,$p=0.022$)。

根据上述检验结果,WTA 和 WTP 均显著高于 VPLP。原因是 WTA 和 WTP 均为百分百泄露概率情境下的个人隐私价值认知,而 VPLP 是在隐私泄露概率为 30% 情境下的个人隐私价值认知。在其他条件不变的情况下,隐私泄露概率由 100% 降低至 30%,用户对其个人隐私价值认知也随之降低。因此,在个人数据保存系统中,应采取严格的隐私保护措施以降低隐私泄露概率从而减轻用户隐私担忧。另外,WTA 显著高于 WTP 的检验结果说明了“隐私悖论”现象的广泛存在,用户关注个人隐私,却不愿意为保护个人隐私付出更多努力。

## 6 结语

本研究测量出受试对象的平均 VPLP 约为 89.5 元,其中,本科生、硕士生和博士生的平均 VPLP 分别约为 83.3 元、91.6 元和 96.8 元。与确定泄露情境相比,泄露概率情境下,用户对其个人数据隐私价值认知显著降低。另外,在 VPLP 实验中,有 2.5% 的用户始终拒绝选择方案 B,一方面可能是因为该部分用户对隐私泄露概率风险的厌恶程度极高,另一方面可能是因为该部分用户对个人隐私的价值认知过高导致其无法接受隐私披露行为,也可能是两者共同作用的结果。

本研究设计的泄露概率隐私价值计量模型对于不同类型个人隐私价值计量具备一定的通用性。不同的企业、机构可根据其使用目的选择不同类型的个人隐私数据进行测量,比如在线营销企业可能更关注用户的行为数据以用于分析用户的购物偏好,而法院更关注容易产生纠纷的个人财务数据等。需要注意的是不同类型个人隐私数据的价值不同,且不同类型个人隐私数据的泄露概率也存在差异,使用该方法时应基于个人隐私数据类型设置合理的实验参数。

本研究的优点:①具有隐含性。采用实验从用户行为中间接推断出泄露概率隐私值,而不是用户直接的报告值。一方面,有研究表明[29]用户的陈述偏好通常不同于观察到的行为(比如“隐私悖论”),从观察到

的用户选择中推断偏好，即使在实验室相对人为的情境中进行实验，也比在一个假定情境中用户的直接报告更接近用户的真实认知；另一方面，隐私价值的直接测量（如WTA、WTP）迫使用户有意识地选择答案，这是不可靠的，因为用户可能难以准确评估与隐私相关的风险和损失，相关研究[30-32]表明针对隐私价值的测量，间接方法优于直接调查。②同时考虑个人隐私内容价值和泄露概率。本研究设计的泄露概率隐私价值计量模型与已有研究的显著区别是，前者呈现隐私泄露的概率风险，后者展示隐私泄露的确定威胁。因隐私泄露遵循随机过程，故隐私决策时个人内容隐私价值和泄露概率均至关重要。现实中，人们必须决定投入多少资金来保护自己的信息免受不确定的随机威胁。确定威胁情境中引发的隐私偏好与随机威胁情境中的行为可能不完全一致，因为后者存在人们的风险厌恶程度。因此，相比于隐私泄露的确定性，隐私披露的随机风险性更符合实际，也更客观，故更具有研究价值。

本研究的局限性主要体现在隐私泄露概率取值和用户样本两方面：①由于缺乏隐私泄露方面的权威统计数据而无法基于隐私泄露的实际情况进行取值，本文基于预实验的结果分布，将隐私泄露概率设置为30%，可能存在一定实验误差；②为了方便调查，本文选取的用户样本均为郑州大学在校学生，用户样本缺乏地域广泛性和人口统计学特征多样性，今后的研究一方面要扩大用户样本的地域覆盖范围以避免地域经济、文化等宏观因素影响，另一方面要扩大用户样本的社会覆盖范围，以分析人口统计学特征（尤其是受教育程度、年龄、职业等个体因素）对VPLP的影响。



**参考文献：**

[1] 新华网. 华住5亿条用户信息疑泄露 警方已介入调查[EB/OL]. [2020-11-12]. http://www.xinhuanet.com/fortune/2018-08/29/c_1123343927.htm.

[2] 搜狐网. LinkedIn遭黑客攻击，约1.59亿用户数据被窃[EB/OL]. [2020-11-12]. https://m.sohu.com/a/292645813_557054.

[3] HANN I H, HUI K L, LEE T, et al. Online information privacy: measuring the cost-benefit tradeoff [C]// Proceedings of the international conference on information systems 2002. Illinois: AIS elibrary, 2002.

[4] ACQUISTI A, GROSSKLAGS J. An online survey experiment on ambiguity and privacy[J]. Communications & strategies, 2012, 1(88):19-39.

[5] POTOGLOU D, PATIL S, GIJON C, et al. The value of personal information online: results from three stated preference discrete choice experiments in the UK [C]// Proceedings of the 21th European conference on information systems. Netherlands: ECIS Press, 2013.

[6] LIM S, WOO J R, LEE J, et al. Consumer valuation of personal information in the age of big data[J]. Journal of the Association for Information Science and Technology, 2018, 69(1): 60-71.

[7] 邓胜利，赵海平. 信息泄露情境下的个人信息价值评估及个体差异：基于离散选择模型的实证研究[J]. 情报学报，2019，38(3):266-276.

[8] GROSSKLAGS J, ACQUISTI A, HEINZ H J. When 25 cents is too much: an experiment on willingness-to-sell and willingness-to-protect personal information[C]// Proceedings of the 6th annual workshop on the economics of information security. Cambridge: WEIS Press, 2007.

[9] 黄逸珺，陆桐，闫强. 电子商务网站个人信息价值评估[J]. 北京邮电大学学报(社会科学版)，2017(5):33-41.

[10] HUBERMAN B A, ADER E, FINE L R. Valuating privacy [J]. IEEE security & privacy, 2005, 3(5): 22-25.

[11] 臧国全，张凯亮，闫励. 个人数据价值计量研究——基于改造的BDM机制[J]. 图书情报工作，2020, 64 (7):103-109.

[12] KRASNOVA H, HILDEBRAND T, GUNTHER O. Investigating the value of privacy in online social networks: conjoint analysis [C]// Proceedings of the international conference on information systems. Phoenix: ICIS Press, 2009.

[13] KIM J, NAM C, KIM S. The economic value of personal information and policy implication [C]// Proceedings of the 26th European regional ITS conference. Los Angeles: ITS Press, 2015.

[14] PU Y, GROSSKLAGS J. Towards a model on the factors influencing social app users' valuation of interdependent privacy [J]. Processing on privacy enhancing technologies, 2016(2): 61-81.

[15] TSAI J Y, EGELMAN S, CRANOR L, et al. The effect of online privacy information on purchasing behavior: an experimental study [J]. Information systems research, 2011, 22(2): 254-268.

[16] BERESFORD A R, KUEBLER D, PREIBUSCH S. Unwillingness to pay for privacy: a field experiment[J]. Economics letters, 2012, 117(1): 25-27.

[17] EGELMAN S, FELT A P, WAGNER D. Choice architecture and smartphone privacy: there's a price for that[M]. The economics of information security and privacy. Heidelberg: Springer, 2013.

[18] KRASNOVA H, ELING N, ABRAMOVA O, et al. Dangers of "Facebook login" for mobile apps: is there a price tag for social information? [C] //Thirty fifth international conference on information systems. Auckland: ICIS Press, 2014.

[19] KIM J E, YEO J. Valuation of consumers' personal information: a south Korean example [J]. Journal of family and economic issues, 2010, 31(3): 297-306.

[20] OTSUKI M, SONEHARA N. Estimating the value of personal information with SNS utility[C]// Proceedings of the eighth international conference on availability, reliability and security. Los

Alamitos: IEEE Computer Society Press, 2013.

[21] DANEZIS G, LEWIS S, ANDERSON R. How much is location privacy worth? [C]//Proceedings of the fourth workshop on the economics of information security. Cambridge: WEIS Press, 2005.

[22] SPIEKERMANN S, KORUNOVSKA J, BAUER C. Psychology of ownership and asset defense: why people value their personal information beyond privacy [C]// Proceedings of the international conference on information systems. New York: ACM Press, 2012.

[23] 百度汉语. 风险[EB/OL]. [2020－11－13]. https://hanyu.baidu.com/zici.

[24] BINSWANGER H P. Attitudes toward risk: experimental measurement in rural India [J]. American journal of agricultural economics, 1980, 62(3):395－407.

[25] BECKER G M, DEGROOT M H, MARSCHAK J. Measuring utility by a single response sequential method[J]. Behavioral science, 1964,9(3):226－232.

[26] LAWRENCE N W. 社会研究方法:定性和定量的取向[M]. 郝大海,译. 北京:中国人民大学出版社, 2000.

[27] KAHNEMAN D, TVERSKY K A. Prospect theory: an analysis of decision under risk [J]. Econometrica, 1979, 47(2):263－291.

[28]陈蓉,王宜峰,邱紫华. 隐含风险厌恶:度量、影响因素与信息含量[J]. 厦门大学学报,2016(1):116－127.

[29] ACQUISTI A, TAYLOR R, WAGMAN L. The economics of privacy [J]. Journal of economic literature, 2016, 54(2): 442－492.

[30] GRAEFF T R, HARMON S. Collecting and using personal data: consumers' awareness and concerns [J] Journal of consumer marketing, 2002, 19(4):302－318.

[31] LEWIS K, KAUFMAN J, CHRISTAKIS N. The taste for privacy: an analysis of college student privacy settings in an online social network [J]. Journal of computer-mediated communication, 2008, 14(1):79－100.

[32] PREIBUSCH S. Guide to measuring privacy concern: review of survey and observational instruments [J]. International journal of human-computer studies, 2013, 71(12):1133－1143.

**作者贡献说明:**
张凯亮:数据采集与分析,图表绘制,撰写初稿;
臧国全:提出研究思路,设计研究方案,修订终稿。



## Measurement of Personal Data Privacy in the Context of Probability of Leakage

Zhang Kailiang[1]  Zang Guoquan[2,3]
[1] School of Politics and Public Administration, Zhengzhou University, Zhengzhou 450001
[2] School of Information Management, Zhengzhou University, Zhengzhou 450001
[3] Research Institute of Data Science, Zhengzhou City, Zhengzhou 450001

Abstract: [**Purpose/significance**] Compared with the situation in which the probability of personal data leakage is 100 percent, the situation in which personal data is leaked with a certain probability is more common. Thus, this paper aims to measure users' personal data privacy value under the certain probability of privacy leakage, which puts forward a new perspective of privacy measurement and the measurement results are of practical significance to privacy classification protection. [**Method/process**] Based on the multiple price list, the user's return on financial risk is measured. Modified the implementation mechanism of multiple price list to elicit users' decisions between the risk-free scheme and the scheme with the probability of privacy leakage. Based on the above steps, value of privacy under leakage probability of users can be measured. [**Result/conclusion**] When the probability of privacy leakage is 30%, users' average VPLP in the social networks is about RMB 89.5; at the same time, when the probability of privacy leakage is 100%, users' average "willing to accept" and "willing to pay" of personal data in the social networks is about RMB 124.1 and RMB 93.8. Users' VPLP depends on the value of personal privacy itself and probability of privacy leakage.

Keywords: personal privacy  leakage probability  privacy value  multiple price list mechanism